

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE Nov, 1, 1999		3. REPORT TYPE AND DATES COVERED Final 07/01/98 - 06/30/99
4. TITLE AND SUBTITLE RETRIEVAL FROM VIDEO AND PICTORIAL DATABASES EMPLOYING SIMILARITY AND MOTION			5. FUNDING NUMBERS G DAAG55-98-1-0424	
6. AUTHOR(S) JEZEKIEL BEN-ARIE, A. PRASAD SISTLA & CLEMENT YU				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF ILLINOIS AT CHICAGO 809 S. Marshfield, Room 502 M/C 551 Chicago, IL 60612-7205			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER <del>P37714-MA</del> 37714.1-C1	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  This project focuses on research towards the development of a system for similarity based retrieval from video and pictorial databases. The specific aims of the work are development of an image understanding system for detecting humans in videos, development of an expressive query language called Hierarchical Temporal Logic (HTL), for spatio-temporal queries on video databases and an initial development of activity recognition module that tracks and recognizes human actions in videos. In this project, we have already developed head and face detection algorithms and also developed an initial tracking system for human body parts. An initial graphical base query system was developed for expressing spatio-temporal queries on video databases. All these algorithms are now combined in a web demo <a href="http://arik.eecs.uic.edu/cgi-bin/vdsearch.cgi">http://arik.eecs.uic.edu/cgi-bin/vdsearch.cgi</a> Our plans for the coming six months which are left in this 1 year program is to extend our tracking, recognition and querying algorithms. We plan to develop more robust head and limb detection and also to develop an initial algorithm for human activity recognition as well as extending the querying system to include hierarchical query structures.				
14. SUBJECT TERMS  Video and Pictorial Databases, Similarity Based Retrieval, Image and Motion Understanding, Human Motion			15. NUMBER OF PAGES 7	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

# University of Illinois at Chicago

**Jezekiel Ben-Arie**  
**Professor**

Department of Electrical and  
Computer Engineering M/C 154

851, S. Morgan St.  
Chicago, IL 60607

Office: (312) 996-2648

Fax: (312) 996-2648

Email: [benarie@ece.uic.edu](mailto:benarie@ece.uic.edu)

**December 19, 2001**

Ms. Bessie B. Oakley  
Chief, Information Control and Analysis office  
Dept. of The Army  
Army Research Office  
P.O.Box 12211  
Research Triangle Park  
NC 27709-2211  
ATTN: AMSRL-RO-BI

**Re: 37714-CI, Agreement Number DAAG559810424**

2001 DEC 31 AM 10:00

Dear Ms. Oakley,  
Enclosed is a copy of the final report we submitted in Nov. 1999 to the ARO. I am sorry  
you can not find it, but I sent it few times also to Dr. Hislop.

Please acknowledge receipt of the final report.

Sincerely Yours,

*J. Ben-Arie*

Jezekiel Ben-Arie.  
Professor.



20020125 290

# **Retrieval from Video and Pictorial Databases Employing Similarity and Motion**

**Jezekiel Ben-Arie, A. Prasad Sistla and Clement Yu**

## **1. List of Manuscripts**

1. Shroff, H.K., and Ben-Arie, J., "Finding Shape Axes Using Magnetic Fields," to appear in **IEEE Transactions on Image Processing**, 1999.
2. J. Ben-Arie and D. Nandy "A Neural Network Approach for Reconstructing Surface Shape from Shading," **IEEE International Conference on Image Processing (ICIP'98)**, Chicago, IL, Vol. II, pp. 972-976, October 1998.
3. Nandy, D. and Ben-Arie, J., "Shape from Recognition and Learning: Recovery of 3-D Face Shapes," **IEEE 1999 Conference on Computer Vision and Pattern Recognition**, Fort Collins, CO, 1999.
4. Wang, Z. and Ben-Arie, J., "Generic Object Detection using Model Based Segmentation," **IEEE 1999 Conference on Computer Vision and Pattern Recognition**, Fort Collins, CO, 1999.

## **2. Scientific Personnel**

1. Prof. Jezekiel Ben-Arie (PI), Department of Electrical Engineering and Computer Science, University of Illinois at Chicago.
2. Prof. A. Prasad Sistla (Co-PI), Department of Electrical Engineering and Computer Science, University of Illinois at Chicago.
3. Prof. Clement Yu (Co-PI), Department of Electrical Engineering and Computer Science, University of Illinois at Chicago.

## **3. Report of Inventions None**

## **4. Scientific Progress and Accomplishments**

### **Introduction**

This project proposes research towards the development of a system for similarity based retrieval from video and pictorial databases. The salient features of the proposed system are :

- An expressive query language, called Hierarchical Temporal Logic (HTL) [14], for expressing spatio-temporal queries on video databases.
- A modular similarity based retrieval system [11] for the temporal part of the query which can be used on top of any suitable picture retrieval system [9] [14].

REPORT DOCUMENTATION PAGE (SF 298)  
(Continuation Sheet)

- An image understanding system for extracting the objects and activities [1] [2] [3] [4] [5] [6] [12] in the videos in the database which uses local and global color characteristics, motion signatures, shape characteristics and texture.
- An activity recognition module that recognizes human actions will be used to interpret video shots with respect to temporal events.
- A dictionary based system for translating the parts of the query into feature vectors [4] [5] [6] [12] [15] which are searched in a space consisting of the feature vectors corresponding to the videos present in the database.
- An incremental learning module which expands the dictionary from examples.

HTL uses the classical temporal operators to specify temporal properties of videos (i.e. the sequencing of video segments), and it employs level modal operators to specify such properties at different levels in the video hierarchy [7] [9]. At the atomic predicate level, the language allows specification of properties on the contents of a single video segment including objects and motions (activities). For a given user query, each atomic predicate in the query is translated into feature vectors [4] [5] [6] (or hyper regions in feature space) using a dictionary based scheme; these feature vectors are then searched using multi-dimensional indexing in the database consisting of the feature vectors of all the video segments in the video database; the result of this is a similarity list containing entries where each entry contains the id of a relevant video segment and a similarity value denoting how closely it satisfies the atomic predicate. Such similarity lists for the atomic predicates can also be obtained by using a picture retrieval system. The similarity lists for the atomic predicates are combined together to obtain a similarity list for the main HTL query. The unique aspect of our feature vector extraction is the combination of the static generic object characterization in pictures [4] [5] [12] [15] and the motion/activity recognition [11] in videos using a unified approach based on multidimensional indexing both in the image domain and in the action domain.

The project proposes extensions to the HTL query language, by incorporating additional temporal operators in order to make it more expressive. A graphical user-interface for specifying HTL queries will be developed. To improve the accuracy of the similarity based retrieval, alternate similarity functions will be investigated. Techniques such as temporal indexing (i.e. indices on the time dimension) will be investigated in order to enhance the performance of the retrieval system. Also, extensions to the currently implemented picture retrieval system are proposed. The project also proposes development of new methods for segmentation of a video stream into shots using edge detection technique based on feature vectors of the frames. Also proposed are novel image segmentation and methods for person detection based on novel approach of model based segmentation and color characteristics. The research proposes generic recognition of objects and activities employing a flexible dictionary approach that translates atomic predicates into feature vectors which are then matched with corresponding features of video frames/shots. The dictionary can represent also generic inanimate or animate objects and activities by hyper-regions in the feature vector space. It is proposed to develop learning techniques which incrementally expand the dictionary with additional entries and generalizes existing entries.

### **Goals, Objectives, and Targeted Activities**

- Extensions To The Hierarchical Temporal Logic (HTL) Language: The HTL language will be extended by introducing new temporal operators. The temporal part of the extended language will be at least as expressive as regular expressions.
- Extensions To The Similarity List Generator: New similarity functions corresponding to the different temporal operators will be introduced.
- User Interface: A user interface for specifying the atomic predicates in the HTL language will be developed.
- Extracting Features From Video And Segmentation:
  - Segmentation of color images based on local and global color and shape characteristics [1] [3] [4] [5] [6].
  - Using Karhunen-Loeve/Principal-Component Expansion for model based segmentation of shapes and persons [2] [4] [12].
  - Action recognition by modeling human body junctions and recognizing their motions for generic actions [11].
- Matching Video With Queries - Generic Objects And Activities:
  - Development of combined segmentation methods for objects that will provide a robust background for reliable feature extraction and for spectral signature extraction for robust recognition.
  - Investigation of neural networks [3] [13] /3-dimensional frequency domain representation for extracting 3D characteristics of surfaces from monocular images for the purpose of pose-invariant 3D object recognition [6] [16] [17].

### **Significant Theoretical Advances**

1. Discovered a novel Volumetric Frequency Representation [6] [16] [17] that encapsulates both the 3D structure of objects along with a continuum of their views. This establishes a new approach in computer vision which can be for many applications such as pose-invariant object/face recognition, shape reconstruction from single images, 3-D motion recovery.
2. Discovered and developed a novel method for segmentation of objects based on their generic models [4] [12]. This method is very useful in detection of objects and persons in cluttered scenes.

### **Significant Experimental Advances**

1. The Similarity list generator for a subclass of HTL queries, called conjunctive queries, has been implemented. Preliminary results are encouraging.
2. We have succeeded in segmentation of human faces using color and frequency characteristics. We have completed a model-based segmentation scheme that detects man-made objects in cluttered scenes [4] [12].

3. A robust scheme for head detection is now in development with successful results for various poses of human head [3] [6]. Also, a robust scheme for tracking human body parts for the purpose of activity recognition is under construction.

## 5. Technology Transfer

1. We provide a web based interface to our preliminary video database. The web demo is available at <http://arik.eecs.uic.edu/cgi-bin/vdsearch.cgi> The database allows the search of video files with content and action keywords. The interface provides the capability for specifying objects and associated actions in a temporal sequence. The temporal depth provided by the menu is 3. Additional simultaneous specifications of objects and actions can also. Currently the simultaneous number of actions that can be specified is also three. A Java based GUI is under development, which will permit unrestricted models of hierarchical and temporal specifications of video content.
2. In related research, Prof. Yu has also developed a metasearch engine to retrieve text documents from multiple databases. A web demo is available at <http://yu.eecs.uic.edu:8080/demo/> New algorithms for retrieving the N most similar documents with respect to a text query from multiple databases have been developed, giving the capability for optimally searching distributed heterogeneous databases. Different databases are optimally ranked during the search depending on statistics generated by their content. It is shown that the documents retrieved from a small subset of databases by this algorithm with respect to a query are essentially the same as those as if all documents are placed in a single site.
3. Human Resources and Training of Personnel: Minlin Deng: Ph. D. student, Tao Hu, Ph.D. student, Zhiqian Wang: Ph.D. student, Dibyendu Nandy: Ph.D. student, Xiao Du, Ph.D. student, R. Venkatasubramanian: M. S. student.
4. Education and curriculum development at all levels. The PI, Prof. Ben-Arie teaches graduate and undergraduate courses on image understanding, image analysis and image processing. His research has influenced and contributed to the contents of these courses. The Co-PI, Prof. Yu teaches graduate and undergraduate courses in Database systems and other areas of Computer Science. His research in heterogeneous and distributed databases has influenced the material of these courses and introduced novel database techniques to new students. The PI, Prof. Sistla also teaches graduate and undergraduate courses in Databases and other areas of Computer Science. His research in HTL and temporal querying mechanisms has also been reflected in the courses taught by him.

## 6. Project References

1. Shroff, H.K., and Ben-Arie, J., "Finding Shape Axes Using Magnetic Fields," to appear in IEEE Transactions on Image Processing, 1999.
2. Nandy D. and Ben-Arie J., "Generalized Feature Extraction using Expansion Matching," IEEE Transactions on Image Processing, Vol. 8, No. 1, pp. 22-33, Jan. 1999.



REPORT DOCUMENTATION PAGE (SF 298)  
(Continuation Sheet)

3. Nandy, D. and Ben-Arie, J., "Shape from Recognition and Learning: Recovery of 3-D Face Shapes," **IEEE 1999 Conference on Computer Vision and Pattern Recognition**, Fort Collins, CO, 1999.
4. Wang, Z. and Ben-Arie, J., "Generic Object Detection using Model Based Segmentation," **IEEE 1999 Conference on Computer Vision and Pattern Recognition**, Fort Collins, CO, 1999.
5. Ben-Arie J. and Wang Z., "Iconic Representation and Recognition Using Affine-Invariant Spectral Signatures," **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 20, No. 6, pp. 604-619, June 1998.
6. Ben-Arie J. and Nandy D., "A Volumetric/Iconic Frequency Domain Representation for Objects with application for Pose Invariant Face Recognition," **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 20, No. 5, pp. 449-457, May 1998.
7. Liu, K., Rishie, Sistla, A. P., Yu, C., "Query Processing in Video Retrieval Systems", **14th Intl Conference on Data Engineering**, Orlando, Florida, March 1998.
8. Sistla, A. P., Yu, C., "Retrieval of Pictures Using Approximate Matching", **Multimedia Databases---Issues and Research Directions**, Edited by V.S. Subrahmanian and Sushil Jajodia, Springer-Verlag 1996.
9. Sistla, A. P., Venkatasubramanian, R. and Yu, C., " Similarity based Retrieval of Videos " **IEEE International Conference on Data Engineering**, April 1997, Birmingham, U.K.
10. Sistla, A. P., Yu, C., et al, "Similarity based Retrieval of Pictures Using Indices on Spatial Relationships", **Proceedings of the 21st VLDB Conference**, Sept 1995, Zurich, Switzerland.
11. Z. Wang and J. Ben-Arie, "3D Motion Estimation Using Expansion Matching and KL Based Canonical Images," **IEEE International Conference on Image Processing (ICIP'98)**, Chicago, IL, October 1998.
12. Z. Wang and J. Ben-Arie, "Model Based Segmentation and Detection of Affine Transformed Shapes in Cluttered Images," **IEEE International Conference on Image Processing (ICIP'98)**, Chicago, IL, October 1998..
13. J. Ben-Arie and D. Nandy "A Neural Network Approach for Reconstructing Surface Shape from Shading," **IEEE International Conference on Image Processing (ICIP'98)**, Chicago, IL, Vol. II, pp. 972-976, October 1998.
14. D. Nandy and J. Ben-Arie, "EXM Eigen Templates for Detecting and Classifying Arbitrary Junctions," **IEEE International Conference on Image Processing (ICIP'98)**, Chicago, IL, October 1998.
15. Wang, Z., and Ben-Arie, J., "SVD and Log-Log Frequency Sampling with Gabor Kernels for Invariant Pictorial Recognition," **IEEE International Conference on Image Processing (ICIP'97)**, Santa Barbara, CA, October 1997, Vol. III pp. 162-165.
16. Nandy, D., and Ben-Arie, J., "Using the Fourier Slice Theorem for Representation of Object Views and Models with application to Face Recognition," **IEEE International Conference on Image Processing (ICIP'97)**, Santa Barbara, CA, October 1997, Vol. III pp.

REPORT DOCUMENTATION PAGE (SF 298)  
(Continuation Sheet)

332-335.

17. Ben-Arie, J. and Nandy, D., "Representation of Objects in a Volumetric Frequency Domain with application to Face Recognition," **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)**, San Juan, Puerto Rico, June 1997, pp. 615-620.
18. Ben-Arie, J. and Wang, Z., "Pictorial Recognition Using Affine-Invariant Spectral Signatures," **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)**, San Juan, Puerto Rico, June, 1997, pp. 34-39.